DOCUMENT RESUME

ED 165 088 CS 004 470

AUTHOR Layton, James R.
TITLE An Analysis of Three Tests: "Durrell Analysis of Reading Difficulty," "Diagnostic Reading Scales" (Spache), and the "Analytical Reading Inventory" (Moe and Woods), Based upon Scores Attained by Elementary Students in Grades One, Two, Four, and Six.
PUB DATE [78]
NOTE 48p.; Report prepared at Southwest Missouri State University

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Comparative Analysis; Elementary Education; Oral Reading; Readability; Readability Formulas; *Reading Diagnosis; Reading Level; Reading Research; *Reading Tests; Silent Reading; *Test Reliability; *Test Selection

ABSTRACT

The distinguishing characteristics and relationships found among the "Analytical Reading Inventory," the "Durrell Analysis of Reading Difficulty," and the Spache "Diagnostic Reading Scales" in terms of readability levels, oral and silent reading results, and independent and instructional reading levels were examined in a study involving 117 first, second, fourth, and sixth graders. Each student was administered the oral and silent selections of the three tests during a three-day period. The results indicated that, of the three tests, the readability levels of the Spache scales were the most commensurate with the numerical levels specified for the reading selections. However, no significant differences were found in the scores on the oral and silent reading subtest levels of the three tests. (Tables of data, organized by grade level, are appended.) (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED165088

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

AN ANALYSIS OF THREE TESTS: DURRELL ANALYSIS OF READING DIFFICULTY, DIAGNOSTIC READING SCALES (SPACHE), AND THE ANALYTICAL READING INVENTORY (MOE AND WOODS), BASED UPON SCORES ATTAINED BY ELEMENTARY STUDENTS IN GRADES ONE, TWO, FOUR, AND SIX

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James R. Layton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

James R. Layton, Associate Professor
Southwest Missouri State University
Springfield, Missouri

- Working Paper -

Introduction

The utilization of informal, individually administered tests by clinicians, special reading teachers, and classroom teachers is a relatively widespread practice. In almost any professional book related to reading education, writers recommend that informal, individually administered tests be used to determine students' reading levels so that appropriate levels of materials may be assigned and also that the results of the tests be used to determine students' strengths and weaknesses in word recognition and comprehension. Many writers recommended also that teachers construct tests and thus, provide themselves with test materials commensurate with the reading materials contained in their classrooms. The results of those procedures supposedly would improve the accuracy of determining the materials that would be too difficult, appropriate, or too easy for members of the class when involved in direct teaching sessions, independent activities, or study activ-

CS004470

ities. Farr (1969a) stated that informal tests were more reliable and valid measures than standardized tests as a result of their usefulness on a wide variety of occasions. However, Farr warned that informal tests could not be used to compare students, one to another or one to a group. The reason posited for that conclusion was that informal, individually administered tests were administered to students under varying conditions and that their validity and reliability were questionable as a result of the inconsistency of administration procedures. Finally though, Farr concluded that if informal reading inventories, based upon materials used in the classroom, yielded reliable estimates of students' performance when reading in instructional materials in the classroom, then the tests would be of value to teachers.

Spache (1976a) presented the most comprehensive review of informal testing and evaluation to date. Following a review of many studies, Spache concluded that selections chosen for individually administered I.R.I.'s would seldom be representative of the remaining portions of the book, story, or chapter from which the selection was extracted. Spache also presented evidence to support his contention that the readability levels of selections and their progressive difficulties from one level to the next would vary from teacher to teacher if the tests were teacher constructed. Spache concluded that standardized individually administered tests were more carefully scaled and more truly representative of materials that students would be assigned to read at varying grade levels. Spache also thought that the grade equivalent intervals between graded selections of standardized, individually administered tests were probably more accurate as a result of controls placed upon their construction. Spache concluded that until the contents of I.R.I.'s were scaled and the selections subjected to standardization that the tests could not be employed to determine the true reading levels of pupils, and that, "they represent . . . reading . . . of unscaled paragraphs of unknown difficulty."

Standardized Individually Administered Tests

Farr (1969, b) separated his treatment of informal tests (I.R.I.) and individual tests which were defined as, "tests that can be administered to only one examinee at a time." Farr's distinction is understandable for many authorities have recommended that informal reading inventories may be used to tests groups if silent reading performance is to be measured. As a result of reviewing the studies of a variety of individual reading tests, Farr concluded that specific individually administered, standardized reading tests could be used in various combinations to confirm improvement or gains in reading of students enrolled in remedial reading classes.

Farr did not view the Durrell Analysis of Reading Difficulty (Durrell, 1955; Durrell test) as a valid measuring device to diagnose specific reading subskills. But based upon several studies reported, Farr concluded that Durrell's test was a valid test of reading achievement and that it was valid for use in diagnosing difficulties students were experiencing in reading when teachers' ratings and standardized tests were used as the criteria. Farr also cited a report from a research team in which they found students' oral reading subtest scores between Durrell's test and the Diagnostic Reading Scales (Spache, 1963; Spache scales) to be highly correlated, but significantly different in terms of grade level mean scores on the oral reading selections. Farr reported that Spache's test yielded mean grade level scores approximately one-half of a grade level above the scores yielded from Durrell's test. Farr reported that Spache's claim that silent reading scores should be superior to oral reading scores was not supported by the researchers cited. They found that only fifty-three percent of the students scored higher on the silent reading paragraphs than on the oral selections.

Farr presented other research findings that were interpreted to mean that as students move beyond second grade oral reading levels, that silent

reading achievement should surpass oral reading achievement, a view harmonious with that of Spache. Farr also posited the idea that errors and habits observed during mature students' oral reading performance could not be used to predict or infer that the same characteristics were operational in silent reading performance. Therefore, according to the researchers cited by Farr, lower correlations should be found between the oral and silent reading scores obtained through individually administered tests as measurements are secured from students for each successively higher grade level. The basis for that claim was that for each successfully higher grade level, oral reading and silent reading abilities would become differentiated and independent, thus should result in students' scoring differently on the two measures.

Spache (1950) drew conclusions in keeping with Farr's that the oral reading subtests of the Analysis of Reading Difficulty and the Diagnostic Reading Scales, used in accordance with specified procedures would yield comparable results. However, Farr did refer to one researcher who criticized all individually administered standardized test designers for attending to the number of errors committed by students in the oral subtests rather than considering the types of errors. That observation will be attended to later in this presentation.

Spache (1976, b) related that no validity or reliability studies were available to support the Durrell Analysis of Reading Disability as a useful instrument to measure reading potential, but otherwise Spache did not review research related to his test or Durrell's test in one of his latest work (Spache, 1976, a; b). However, in Spache's (1976, c) second most recent book, he viewed Durrell's first five selections in oral and silent reading to be similar in structure and difficulty but totally lacking in their relationship to materials used by students in the first five grades. Additionally when applying a readability estimate (Spache's own) to both sets of selections,

Spache reported that the fifth level selections were judged to be lower third grade levels; fourth level selections, upper second grade levels; third level selections, lower second grade levels; second level selections, upper first grade levels; and first level selections, mid-first grade levels. One of Spache's strongest criticisms of the Durrell test was directed toward the use of rate norms in both the silent and oral reading selections to determine students' reading levels. Spache concluded that grade level scores derived from using Durrell's test would tend to be lower estimates of reading ability when compared to other standardized reading tests.

Spache (1976, c) reviewed the Diagnostic Reading Scales and presented numerous authors' views of the weaknesses within the test. For each criticism, Spache cited the results of statistical studies to defend the philosophical base upon which the test was constructed as well as its use to measure true gains students achieved as a result of remedial instruction. Spache stated that one of the strengths of the Diagnostic Reading Scales was in the standardization procedures for determining oral reading level. The norms for oral reading errors were established by determining students' mean scores then adding one standard deviation. As a result, students' scores on the Spache scales oral reading levels should be higher than on other tests. But, according to Spache, the procedure of basing reading level estimates on the errors students commit, rather than arbitrary standards determined by the designers of other tests, is the only scientifically accepted method for making those determinations. Spache revealed that the Spache scales were both valid and reliable for measuring students' reading levels and determining growth as a result of instruction. However, he admitted that the levels obtained would not be readily useful for choosing materials. No tests, according to Spache, have been sufficiently developed to allow that. But Spache viewed that as not a weakness in test design, but a result of variability in

the difficulty levels of materials in the classroom and the inability of average teachers to evaluate, analyze, and classify materials to allow for test results/instructional materials matching.

Commercially Prepared I.R.I.'s

Anderson (1977, a) reviewed research studies in which the characteristics and/or values of commercial, informal reading inventories (I.R.I.) were investigated. Anderson viewed the commercial products as being somewhat less than informal as a result of their being nonlocally produced and not based upon instructional materials found in classrooms. Nevertheless, Anderson conceded that the use of the tests to estimate reading levels quickly and to determine characteristics of the reading processes employed by individual students was possible. He commented that commercially prepared tests usually retain the characteristics of informal reading inventories.

Anderson's comments were harmonious with Spache's and Farr's when he viewed the most serious problem of the commercially prepared tests to be their lack of reliability and validity. Anderson believed that if systematic or recurring use of a commercially prepared I.R.I., not related to instructional materials, was to be employed then the tests must be valid and reliable or internally consistent. Anderson also viewed the standardization or consistency of the administration of commercially prepared I.R.I.'s among examiners to be of paramount importance if test results were to be of value.

In his comparative review of tests, Anderson (1977, b) was critical of Woods and Moe's (1977) Analytical Reading Inventory (A.R.I.) as a result of the lack of validation studies presented. Although the designers did, according to Anderson, make an attempt to present data intended to validate the instrument, which was more than other designers offered. Anderson was also critical of Woods and Moe for not presenting reliability data except for the equivalency of alternate forms.

One characteristic of tests alluded to by Anderson, but not by Spache nor Farr in their analyses of the Durrell test or the Spache scales, was the matter of passage dependency, or the necessity of an examinee to read a selection to answer the comprehension questions. Anderson cited several studies from which reports were issued that many students can answer a significant number of comprehension questions accompanying reading selections on tests without reading the material. In Anderson's estimation, scores yielded from that situation would be spurious and confound the process of obtaining reliable and valid scores. Questions that measure experiential backgrounds of students should be eliminated from I.R.I.'s according to Anderson. However, he did not find evidence of those types of questions in the A.R.I. In fact, questions on the A.R.I. were judged as, "well-constructed, carefully devised, . . . passage dependent . . ."

Anderson concluded that the A.R.I. was useful as a diagnostic tool and that by using it, quantitative and qualitative assessment of students' reading performance could be determined. Examiners who employ the test are directed to analyze errors in accordance with Goodman's (1973) miscue analysis technique (oral miscues) in which he essentially provided guidelines for examiners to differentiate between meaning and non-meaning errors occuring in students' oral reading performance. Earlier in this presentation, one of the criticisms directed toward individually administered oral reading tests was the insensitivity of tests designers to the students' repetion of the same types of errors (repititions, miscalling articals, and substituting one word for another word consistently) that did not interfere with comprehension, but would depress the oral reading score as a result of the number of errors committed. Spache made provisions for those types of errors; Durrell did not.

Conclusion: The Tests, Likenesses and Differences

The reviews of the literature and subsequent interpretations by Farr, Spache, and Anderson lead to the conclusion that the three tests being con-

sidered (Analytical Reading Inventory, Diagnostic Reading Scales, and Durrell Analysis of Difficulty) are very similar in many respects but that they are also very different. All three were cited as useful and valid when employed to determine students' reading levels. However, only the Spache scales were purported to have been subjected to statistical validity and reliability tests.

Oral Reading

The three tests contain oral reading passages for students to read. Examiners are directed to record word recognition errors and comprehension errors and use those errors to determine the level at which testing was to be concluded. The A.R.I. designers used a combination of the Betts' criteria and Goodman's miscue analysis techniques reading levels. Spache collected data from students' performances and developed a statistically supported scheme (mean plus one standard deviation) to determine the ceiling selection for word recognition in oral reading and sixty percent accuracy in comprehension questions. He considered that to be the instructional level. Durrell directed examiners to cease oral reading testing when a student committed either more than seven word recognition errors or missed more than one comprehension question. However, the reading levels on the Durrell test are determined by computing the median score of the grade equivalents from the selections read. The grade equivalents are determined solely on rate. Durrell did not direct examiners to determine an instructional or independent reading level, therefore comparisons of the results of students' scores cannot be classified as instructional or independent except by an arbitrary system developed by the examiner.

Silent Reading

Woods and Moe did not provide a systematic plan for evaluating students' silent reading abilities. Their only directions were in the form of a sug-

gestion that silent reading exercises be used with "older" students who the examiner thinks could comprehend better silently than orally. The examiner is directed to administer a silent selection from an alternate form at the same level the student read orally and then compare the students' performance in comprehension between the two selection. No standards or criteria for analysis are presented. That recommendation is somewhat harmonious with Durrell's directions. Once oral reading is concluded on the Durrell test, the examiner directs the student to read silently, selections that are at the same levels as those read ora-ly. Following each selection the student recites all the information that can be recalled (unaided). Once the supply of memories is exhausted, the examiner asks questions to determine if the student can recall more information (aided). The unaided and aided memories are combined and used along with the number of seconds required by the examinee to read the selection to determine the level of silent reading for each selection. The resultant grade levels on each selection are ordered and the median score (grade equivalent) is computed. Durrell (1940) did indicate that a student should score seventy-five percent or higher on combined memories on selections thought to be suited to the students' reading level. Again, Durrell did not provide directions for determing independent or instructional levels.

Spache designated the silent reading level derived from the Spache scales as the independent level. Although the scope of this presentation does not allow the inclusions of his reasons, Spache considered silent reading skills to represent a student's optimum reading ability when working independently and reported evidence to support his claims. Spache directs examiners to conclude silent reading exercises when students score below sixty percent correct on the comprehension questions, which is the same standard used for oral reading. Spache derived those standards from comparisons of students' mental ages to their reading scores. Although the designers of the three

tests gave directions for administering oral and silent reading subtests, there are few specific likenesses in either administration, interpretation, scoring for errors, or determining students reading levels among the directions.

Readability

The A.R.I. reading selections were subjected to readability analysis using the Spache readability formula and the Harris-Jacobson Formula 2. Anderson (1977, b) was complimentary of the thoroughness of their actions. The Spache scales were subjected to analysis through use of the Spache and Dale-Chall readability formulae. Spache also used teachers' judgment to aid in determining the difficult levels of his selections. The corresponding levels on the various forms of both the A.R.I. and Spache scales were subjected to readability and reliability studies and determined to be parallel, reliable, and interchangeable. However, that statement refers to selections within each test, not between the two tests. Durrell was criticized by Spache (1976, c) for not attending to the readability levels of his selections and also for ignoring to upgrade them since 1937. Spache applied his readability formula to the first five paragraphs of the oral and silent reading selections contained in the Durrell test and reported that the readability levels between the two subtests were parallel, but did not represent the levels indicated by their numerical designations. However, it may be noticed that Durrell numbered his selections and referred to the selections as graded. But, nowhere did Durrell specify that selection one was written at the first grade level, level two at the second grade, and so on. In fact, Durrell referred to the graded selections as appropriate for evaluating students' reading abilities in elementary school but not beyond. Durrell published eight paragraphs, not six. There is proximity of the levels between subtests, but not an upward distinction between levels within the subtests. Grade levels are determined

by computing a median score from several paragraphs rather than one. All that leads to the conclusion that Durrell developed an altogether different system for determining reading levels and analyzing reading behaviors than was developed prior to 1937 or that has been developed since. But the evidence is clear that variability exists among the readability levels of the selections appearing on the A.R.I., Durrell test, and Spache scales. Data to support that claim will be displayed later in this presentation.

The Problem

Selecting a valid and reliable individually administered test to use in special reading classes, clinics, and in summer reading programs is not a simple or easy task. The results of the tests must be useful in guiding the examiner to determine reasonable reading levels for students. The test results should also be useful in determining reading levels that can be distinguished from the levels above and below the level scored. That knowledge will aid the examiner in choosing materials if the materials have been carefully graded also. Another reason that an individually administered test must be reliable in measures of reading levels is to allow examiners to measure gains made during the course of instruction, especially if the time periods are short as are typical of summer reading programs. Tests that are designed to yield results that are insensitive to students' growth following several weeks of daily, intensive instruction may be the source of teacher and student frustration and bewilderment. Lastly, the knowledge that wide variability exists among the results obtained from different tests should lead reading educators to analyze and study tests in a systematic fashion to determine strengths and weakness, likenesses and differences, and applicability to various evaluative and instructional situations. Armed with that knowledge, reading personnel should be able to: (1) choose tests more appropriately to achieve the purposes of administering the tests to students; (2) be able to

adjust scores to reflect students' reading levels that would be commensurate with classroom or clinical reading materials or choose materials at the level of scores obtained; and (3) be able to interchange tests, while being fully aware of the variability between them and be able to adjust results appropriately.

Statement of the Problem

The problem to be solved in this study was to specify the distinguishing characteristic and relationships found among the Analytical Reading Inventory, the Durrell Analysis of Reading Difficulty, and the Diagnostic Reading Scales in terms of readability levels, oral and silent reading results, and independent and instructional reading levels.

Questions to Be Answered

To explore the problem posited, several questions were developed by the investigator to use as guides in pursuing the possible solution to the problem. Those questions were:

1. What variability exists among the readability levels of the graded selections contained in each of the three tests when readability scales are employed to determine the levels and what conclusions may be drawn from the comparisons?
2. What variability exists among the readability levels of the graded selections of the three tests when comparison are made across the tests for each level and what conclusions may be drawn from the comparisons?
3. What difference exists between the students' grade placement, general expectancy levels, and their mean scores on the tests and what conclusions may be drawn?
4. Based upon scores obtained from students in grades one, two, four, and six, what relationships exist between the students' scores on the levels

across the tests, and conclusions may be drawn?

Delimitations

Three common features were identified among the A.R.I., Durrell test, and Spache scales: (1) readability, (2) oral reading, (3) silent reading. Within those three areas, only the studnets' oral reading performance was obtained for the A.R.I. as a result of the authors' directions. The A.R.I. oral reading was classified as: (1) instructional level and independent level, except for first grade students. The Durrell subtests of oral and silent reading were administered to all students, except for first grade students who were excluded from silent reading tasks; the same was true for the Spache scales. The oral and silent subtests of the Spache scales were administered to the students in grades two, four, and six. Woods and Moe did not design the A.R.I. to be used with first grade students. However, the students tested in all grades were in the ninth month of school and an arbitrary decision was made to test the first grade students' oral abilities with all three instruments.

Students selected for participation in the study had recorded intelligence scores between ninety and one-hundred ten. They were all identified by their teachers as being able to perform the assigned tasks in reading and subject matter areas in daily classroom activities. Therefore intelligence test scores nor achievement scores were collected or reviewed. Analysis of or inclusion of that data is beyond the scope of this study. This investigation was conducted primarily to study the characteristics of three individually administered reading tests based upon scores obtained from students who were classified into the broad range of normal in daily classroom tasks and had no apparent problems in intellectual functioning.

Only Form A of the Analytical Reading Inventory was administered to the students who participated in this study. Time limits within which the testing were completed were not sufficient for testing students with all three forms.

Additionally, such intense testing may have depressed the students' enthusiasm for the testing. Only the A paragraphs were used from the Spache scales.

Procedures

Students

The students chosen for participation in this study came from two populations: (1) the university laboratory school and (2) the public schools in the immediate area. Students were not randomly selected and no records were kept of their sex, race, or cultural backgrounds. The students met two criteria: (1) their teacher's judgement that they performed well in the instructional materials and other language/reading activities in the classroom and (2) recorded data that their intellectual levels were within the broad limits of average (90-110) or above. This study was designed to evaluate the characteristics of the three tests cited; therefore, students who were known to have any debilitating reading or intellectual characteristics were not chosen to participate. All of the students were completing the last weeks of the ninth month of the grade in which they were placed.

Schedule

The students were tested during a three week period. Each was administered the oral and silent selections of the A.R.I., Durrell test, and Spache scales during a three-day period. The word lists from the A.R.I. were used according to the designers' directions to find approximate levels to begin testing students. Following the administration of the A.R.I., the examiners tested the students using the oral and silent subtests on the Spache scales and Durrell test, but no other subtests were administered. The subtests were administered and scored in strict accordance with the designers' directions and/or suggestions.

Examiners

Five examiners, competent in the administration of the tests, were employed. Prior to conducting the tests, the examiners met to ascertain that all could follow the same procedures. They also met periodically during the testing period to ascertain that their testing procedures were still commensurate. As a result of the simplicity of the administration procedures used in the oral and silent reading subtests, no problems were encountered.

Scoring and Assigning Grade Equivalents

For each test administered grade levels were assigned in accordance with the designers' directions. The A.R.I. scores were always whole numbers. On the Spache scales, grade level designations were given in half-month scales (4.5, 5.5, etcetera) for grades four and above. On the lower levels scores were assigned at the upper and lower ratings specified by Spache. Matters were not as easy when assigning reading levels to scores derived from the Durrell test. For each grade level, Durrell listed three levels: (1) low, (2) middle, and (3) high. For purposes of statistical analysis, arbitrary assignments of .3 for low, .5 for middle, and .8 for high, were assigned.

Results from Spache's readability scale and the Dale-Chall formula are given hundreths for each reading level. To make comparisons of the readability levels of the selections somewhat easier, the results of the Fry scale were converted to tenths by determining if the coordinates crossed at the lower, middle, or higher portion of the range for each grade level. The same system employed with the Durrell test was used (.3 low; .5, middle; .8 high). Should the reader view that procedure as improper, the tenths designation may be removed and the whole number designation found in the analysis will represent the Fry level.

Statistical Treatment

Means, standard deviations, and standard errors of the means were com-

puted for the students' scores on each subtest. The subtest scores for the students (grades one, two, four, and six) were then subjected to test analysis to determine if differences existed. Pearson product-moment correlations were computed to determine if relationships existed. The independent and instructional reading levels of the students in grades two, four and six were compared and correlated to their silent and oral reading levels on the Durrell test and the Spache scales. The students' oral and silent reading levels on the Durrell test and the Spache scales were directly compared and correlated, but silent levels were not cross compared or correlated to oral on those two tests. The reading levels of the first grade students were treated similarly as described above except that only oral reading measures were obtained and only the instructional level was determined for the A.R.I.

Results and Conclusions

The first two questions posited for investigation were: (1) "What variability exists among the readability levels of the graded selections contained in each of the three tests when readability scales are employed to determine the levels, and what conclusions can be drawn?" and (2) "What variability exists among the readability levels of the graded selections of the three tests when comparisons are made across the tests for each level and what conclusions may be drawn from the comparisons?"

Readability

Table 1 is a display of the readability levels of A.R.I., Durrell test, and Spache scales determined through the use of the readability scales specified. Except for the primer level and first grade level of the A.R.I. which

---

Insert Table 1 about here

---

appear to be relatively the same, levels one through four appear to be commensurate with the designated levels. Beginning with grade five, the A.R.I. levels appear to be one grade higher than the indicated grade level. The grade seven selection is the same as the results determined by using the Harris-Jacobson Formula Z. The levels for selection eight and nine are confusing; a designation of 8+ indicates a level higher than eighth grade level, but what level? Is the examiner to arbitrarily decide or employ another scale to determine those levels? According to Woods and Moe's report, only five of their selections are in proximity of the grade levels for which they were designed.

Reanalysis of the selections by the Fry scale, which was not designed to allow an investigator to compare the number of difficult words through word matching techniques, revealed that only three of the selections were in proximity of their numerical designation. However, selection two contained many one-word sentences and that resulted in reduction in readability when the Fry scale was employed. Viewed simultaneously, based upon the Fry scale results, the readability levels of the selections in Form A of the A.R.I. appear rather spurious. Viewed simultaneously based upon the results from the Spache formula and Harris-Jacobson formula, the same conclusion may be drawn.

The reading levels of the selections in Form A of the A.R.I. do not appear to be sufficiently stable to insure an examiner that a student will be reading selections that are decidedly one level higher as the student progresses through the test. Neither can an examiner depend upon using portions of the test for specific purposes. According to the results obtained from the readability formulas, students reading selections five through nine may read between the fifth grade level or somewhere above the eighth grade level, but who will know? The test is also not applicable to students in grade one; therefore, based upon readability levels, students who are able to read between levels two, three, and four may stand a better chance of being accurately

diagnosed than others.

One concluding comment appears in order. The relative instability of readability formulas is well-known. Many characteristics of the complexity of printed materials cannot be quickly and easily ascertained. However, by using readability formulas and additional suggestions (Bormuth, 1968) for developing materials for students, it appears that test materials could be shaped to conform to certain standards. When the test materials are stabilized, special teachers of reading and clinicians would be able to analyze students' reading based on the nature of the test material as well as determine levels with more confidence.

Spache's observation that the selections of the Durrell test did not appear to become progressively more difficult was true. Analysis by using the Spache and Fry formulae indicated that through the grade five selections, a student will not be reading selections above a third grade level. Beyond those levels, the grade levels obtained through use of the formulae were not commensurate with their numerical designations, They extended into what was considered to be college levels (12.+). However, the Durrell test selections were in proximity between the oral and silent subtest selections. That finding leads to the question as to whether Durrell intended the selections to be progressively more difficult. Computing the median of a student's grade level scores, derived from speed (oral subtest) and speed and recall (silent subtest), is a different procedure from those dictated by Spache or Woods and Moe. The results of a student's score on the Durrell is not an indication of the student's ability to read material written at a specific grade level, but rather a comparison of that student's ability to other students in the norm group, to read selections at a certain speed. Duly considered, the procedure is not unlike the procedures used in group standardized reading tests, except that students set the time factor, which in turn does not deflate or inflate a student's grade level score.

It is apparent that the selections in the Durrell test do not represent their numerical designations. Therefore, they cannot be utilized by examiners if progressively more difficult reading selections are needed. However, it may be noticed that the levels below the sixth selection are written below those levels. The majority are second or third grade level, or just beyond the point at which students should have mastered the basic or sight vocabulary words. The levels of the selections are also stable in sentence length. The proximity between the oral and silent selections may also support the conclusion that Durrell was more aware of the difficulty levels of the selections than reviewers thought. Perhaps Durrell wished to determine the degree of proficiency a student would exhibit when reading materials composed largely of basic sight words and only a few difficult words when compared to readers in other grades. Many authorities cited speed of perception of sight words and speed and ability to understand phrases as two characteristics of students who read well. Perhaps Durrell's procedures are more logical than some writers believed when considered in relationship to what is known or thought to be true about the reading process.

The Spache scales was the only test of the three that contained carefully shaped reading selections when the Spache and Dale-Chall formulae were applied. However, when the Fry formula was applied differences in the levels were identified. On the A selections the results of the Fry scale revealed only one selection to be in proximity of the numerical designation, three of the B selections were in proximity, two of the C selections were in proximity, and only one of the D selections was in proximity. Based on the results of the analysis using Fry's scale, the Spach scale selections do not appear to be any more stable than those of the Durrell Test or the A.R.I. Based on the readability standards used to design the test, the Spache selections appear to be more carefully shaped than those of the A.R.I. Analysis of contents of the selections was beyond the scope of this study.

Conclusion: Readability

Based upon the data presented it appears unlikely that students scores on the three tests would be similar. Considered simultaneously; using the results of all the readability scales, the first level selections displayed in Table 1 ranged from .8 to 2.3; the range for the level two selections was between .8 and 3.3; the range for level four selections is between 1.5 and 6.8; and for level six selections the range is 3.3 to 8.5. The variability in the readability levels, coupled with the three separate standards for administering the tests and determing reading levels leads to the conclusion that the results of students' scores on the three tests would probably be very different. Additionally it appears unlikely that any significant positive relationships would result among students' scores on the oral and silent grade level results. Initial conclusions were also posited that students in the ninth month of school in grades one, two, four, and six would not have mean scores commensurate with the general expected reading levels for those groups.

Grade Placement, General Expectancy, and Mean Scores

The students who participated in this study were in grades one (N=20), two (N=21), four (N=44), and six (N=32). All were considered by their teachers to be able to perform adequately in reading and reading-related tasks in the classroom. A check of the students' records was used to confirm that the students' intelligence scores were well within normal or average ranges (90 to 110) or slightly above. The inclusion of the variables of intelligence and expectancy, were beyond the scope of this study. However, the students participating in this study should have been able to read between the lower levels of the grade into which they were placed or at the next higher level.

Table 2 displays the means, standard deviations and standard errors of the means for the students on the subtests of the A.R.I., Durrell test, and

Insert Table 2

Spache scales. When the data were considered simultaneously for all grades and all subtests, the following conclusions were drawn:

1. The Durrell test scores were on or slightly above students' grade placement in every instance. Oral reading means were almost identical for second grade students and slightly higher for students in grades four and six.
2. The Spache scale scores were on or above grade placement in three instances, one of which was in the higher level of grade two (2.79) for the first grade students. Oral and silent reading means were similar for students in grades two, four, and six.
3. The A.R.I. levels were on or above grade placement in three instances, but using two standards for oral reading confuses the analysis. Only two levels (grades two and four) were commensurate at the instructional level, with the second level being low (2.04). When the standards were higher (independent level) the levels for the sixth grade students were in proximity with their grade placement.

The mean scores obtained by using the Durrell test were more commensurate with the students' grade placement than the scores of the other two tests. However, comparisons between the levels obtained from use of the Spache scales with the students' grade placement were relatively consistent. Mean scores for students in grades one and two were 1.89 and 1.65, respectively, higher than their grade placement. Possibly, the Spache scales levels may be accounting better for expectancy levels than the Durrell test, but that determination cannot be made from the data collected and is beyond the scope of this study.

The A.R.I. scores were inconsistent at the instructional level. Mean scores for grades one and two were much lower than the students' grade placement. Scores for fourth grade students were almost identical to grade placement and the mean score for the sixth graders (8.31) was much higher than grade placement. As would be expected, the independent level mean scores for second and fourth grade students were far below grade placement. Results of the A.R.I. scores do not appear to be as consistent at the independent or instructional level as the mean score results of the Durrell test or Spache scales.

Comparing the students' mean scores between the levels within each test, further comparisons were possible. Table 3 displays the distance between the mean levels of the students at each grade level. The levels of

---

Insert Table 3

---

the Spache scales in oral and silent reading approach a two year difference ($\pm$) between grades six and four, four and two, and a one-year difference between grades one and two. Seemingly the results of the Spache scales scores differentiated between the grades better than the other two tests. Both the Durrell test and the Spache scales yielded high scores for the first grade students (2.56 and 2.79 respectively). The mean scores were an indication that the students could read well. The A.R.I. mean scores did not reflect that ability at the instructional level (.50). The independent level scores for the students would have been lower should it have been computed.

In one instance (between grades two and four) results of the Durrell test approached a two year difference (1.78). Other differences did not

approach the two-year span as closely as the Spache scale results. Except for the 1.78 difference cited, Durrell test results approached a one and one-half year difference between grades four and six and two and four, and a one-half year difference between the students' scores between grades one and two.

The differences between the students' mean scores in grades four and six on the A.R.I. approached three years on the independent level and was 3.31 years on the instructional level. The differences between scores between grades two and four were two and one-half years on the independent level and almost three years at the instructional level. The difference between the first and second grade students' mean scores was 1.54 years. A finding that supports Woods and Moe's contention that the A.R.I. is not suited for first grade students. The levels obtained through using the A.R.I. were puzzling. Woods and Moe's attention to the linguistic principles of Goodman and the subsequent scoring systems should have resulted in higher reading in one sense, but their allegiance to the Betts criteria should have counterbalanced the scoring. But to result in very low scores for students at the end of grade one and a one and one-half year increase for students in grade two is confusing; especially when compared to the results obtained from the Durrell test and the Spache scales. Data collected for this study cannot be used to determine the reasons for the differences. Whether the Durrell test or Spache scales means represented true reading levels or not, the consistency between grade levels was reassuring.

Although a detailed analysis of the standard deviations and standard errors of the means will not be presented, the standard errors of means displayed in Table 2 may be interpreted to mean that the students' true scores on the Durrell test were less likely to fluctuate from the mean further than the Spache scales scores. The Spache scale standard errors of

the means were usually higher than those of the Durrell test. The standard errors of the means for the A.R.I. scores were similar to those of the Spache sclaes; both were slightly higher than those of the Durrell test standard errors of the means which were in the .10's and middle to lower .20's. The range of fluctuation of the students' true scores on the Durrell test may have ranged between .48 and .56 of a year; an average of .38. The range of students' true scores on the various Spache scales subtests may have ranged between .32 and .72 years; an average of .56. The range of the students' true scores on the A.R.I. may have fluctuated .32 and .96 years; an average of .64. Based on the standard errors of means presented, the use of all three tests for the students tested may have resulted in true scores that would not have fluctuated less than 3.12 months or not more than 6.72 months above and below the students' observed scores.

Conclusions: Grade placements, General Expectancy, and Mean Scores

The purpose for conducting this study was to determine likenesses and difference that occurred on the subtests of three individually administered tests when students' scores on the subtests were analyzed. The analysis of the data and subsequent conclusions should be directed toward the tests, with other matters ignored. Therefore, any conclusions directed toward students for whom these tests may be more applicable or the relationship of the scores obtained to instructional materials contained in the classroom will be omitted. Based on the students' scores reflected in the data presented in Table 2 and 3 and the analysis just presented, the following conclusions appear warranted.

1. Except for first grade students, the Durrell test mean scores appeared to be closer to the students' grade placements than the mean scores of the other two tests. The sizes of standard deviations and standard errors of the means are interpreted to mean that the Durrell test results were reliable in distinguishing between grade levels between second,

fourth, and sixth graders. Furthermore, the small standard errors of the means indicated that the chance for fluctuation of students' scores was less than that of the other two tests.

2. The students' grade level means obtained by using the Spache scales were higher for each grade than the students' grade placement. However, that occurence cannot be interpreted to mean that the Spache scales were less accurate than the Durrell test means; they were higher therefore different. The Spache scale means appeared to differentiate between the scores of the first and second grade students slightly better than the Durrell test means and were slightly better than the Durrell test means in closely approximating a two-year difference between the mean scores of students in grades two and four, and four and six. Although the standard deviations and standard errors of measurements for each grade were slightly larger than those of the Durrell test means, the participating students' true scores may have fluctuated slightly further from the observed scores than they may have from the Durrell test scores. Therefore, it appears that the Spache Scales was a reliable test for discriminating between the grade level scores of students in grades one, two, four, and six.
3. Comparisons of the students' mean scores on the A.R.I. between grades was higher in variability than the contrasted means from the other two tests. According to the test results obtained for the oral (instructional level), the first grade students could barely read (.5). There was a 1.54 difference between the means of the first and second grade students. Almost a three year difference (2.99) between the means for second and fourth grade students, and a 3.31 span between fourth and sixth grade students. The A.R.I. mean independent levels between second and fourth graders was 2.47 which was closer to the two-year chronological difference. A difference of almost three years

(2.90) existed between the sixth and fourth grade students' means. In the sixth to fourth grade comparison of the independent level, the standard error of the mean was .48 indicating that those scores could fluctuate as much as one year. The other standard errors of the means were similar to those computed for the Spache scales (.16 - 36).

Based upon mean scores, standard deviations, and standard errors of measurements the Durrell test appeared to give better overall results for second, fourth, and sixth graders when compared to actual grade placement. The Spache scales was superior in distinguishing between grade levels for all grades but higher grade level equivalents were obtained. The standard errors of the means on the Spache scales were slightly higher overall than those of the Durrell test. The results of the scores on both tests were interpreted to mean that the tests were reliable to differentiate between the students' abilities in the population studied.

Results of the scores obtained through the use of the A.R.I. were interpreted to mean that the test results were not commensurate with the chronological time differences between the students' grade placements. However, the standard error of measurement in almost every instance was no greater than those obtained from the Spache scales. If the assumption can be made that students should gain approximately one year's growth for every year in school, then the Durrell test and Spache scales appeared to show that difference. If it can be assumed that a randomly selected group of students should be reading relatively close to grade placement, then the Durrell test results indicated that.

RELATIONSHIPS AND DIFFERENCES BETWEEN SCORES ON MEASURES OF SILENT AND ORAL READING

The fourth question posited for investigating the characteristics of the oral and silent reading subtests of the A.R.I., Spache scales, and Durrell

test was: "Based upon scores obtained from students in grades one, two, four, and six, what relationships existed between the students' scores or levels between the subtests of the tests; what differences existed between the students' scores on the subtests of the tests, and what conclusion may be drawn?" The first part of question four was an inquiry into the relationship that may have existed between various subtests of the A.R.I., Durrell test, and Spache scales. Table four is a display of all correlations computed.

Insert Table four

The correlations between the A.R.I., Durrell Test, and Spache Scales were statistically significant in all instances. The high correlations obtained are interpreted to mean that the students scored relatively the same on any given subtest of one test as they did on the other; also that the subtests scores either accounted for the same things or that the students read and comprehended in the same manner in oral or silent reading whether it was onsidered the inde endent or instructional reading level. It may be noted that the correlations between the Spache scales (silent-independent level) or the Durrell test (silent) and the A.R.I. (independent level) were somewhat depressed at each grade level. There is a distinct possibility that the standards recommended for determining the independent level may change a students' score, thus his rank in the group, and account for the lowered score. The possibility that a difference in oral and silent reading performance could account for the depression was abandoned as a result of the almost perfect correlations between instructional levels on the A.R.I. and silent reading levels on both the Spache scales and Durrell test. Apparently, the students, at all grade levels, who participated in this study read silently very similarly to the manner in which they read orally.

Differences

To answer the second part of question two, correlated t tests were computed between the students' oral and silent subtest levels of the three tests. Table 5 displays the instances in which no statistical differences were found.

Insert Table 5

The scores for sixth grade students (N=32) on the A.R.I. (independent-oral) were very similar to their same scores on the Spache and Durrell oral and silent subtests. However, those are the same subtests that were identified as tending to be less correlated than the other subtests. The A.R.I. (instructional-oral) levels for the sixth grade students were very similar to their silent and oral scores on the Spache scales. Apparently, the scores for sixth grade students on all three tests were significanly related and not significantly different. However the t's of 2.22 and 2.24 between the A.R.I. independent level and Spache scales are approaching the critical value of 2.64 and may not be interpretable as strongly similar for educational purposes. But the sixth graders' scores on the A.R.I. (independent) were very similar to their scores on the other two tests' oral and silent subtests. Also, the sixth grade students' A.R.I. instructional level scores were very similar to their oral and silent reading scores on the Spache scales.

Fourth grade students' (N=44) scores were significantly different in all instances except four. The scores on the A.R.I. instructional level were almost identical to those of the Durrell test (oral). Other nonsignificant differences between the subtests for the A.R.I. and the subtests of the Durrell test and Spache scales occurred between A.R.I. (instructional) and Spache scales (silent). The t test result was 2.50. The fourth graders' A.R.I. independent levels were not significantly different from the Durrell

test's oral and silent subtests (t's = 2.51 and 2.17 respectively). In all three instances the t's were near the critical value of 2.62 (p.01) and may not be interpreted to be significantly similar for educational purposes. The instructional level scores on the A.R.I. and the oral subtests of the Durrell were the only scores of fourth grade students that were very similar.

Scores of the second grade students on all three tests were not significantly different in four instances. The two nonsignificant differences found between the scores on oral and silent subtests of the Durrell test and the A.R.I. instructional level (t's = 2.39 and 2.36, respectively) are near the critical value of 2.70 (p.01) and are not interpretably close. However, the Spache scales (oral) and the Durrell test (oral) scores (t = .996) as well as the silent reading scores (t = .730) were very similar for second grade students. The oral reading subtest scores for first grade students were also similar (t = 1.04). But, that was the only instance in which no significant differences resulted for first grade students. mean scores.

For the population participating in this study there were few instances when their mean subtest scores on the three tests were not significantly different. Only the sixth grade students' scores between the instructional level (oral) and the Spache scales (silent) and the A.R.I. independent level (oral) and the Durrell test and Spache scales (oral and silent) were similar. As a result of the high t values obtained, a few of subtest mean scores were not considered similar enough for educational purposes. Result of the standard errors of the means alluded to earlier in the presentation were also used to support that conclusion.

CONCLUSIONS

Based upon the findings presented in this report, the following conclusions are drawn.

Readability

Based upon the data supplied by the designers of the tests, the readability levels of the Spache scales are more commensurate with the numerical levels specified for the reading selections. The arithmetical differences between the levels are also more consistent than those of the other two tests. The A.R.I. paragraphs are commensurate with the numerical designations of the selections and somewhat consistent differences for the lower level selections. But, the readability levels begin to outdistance the numerical designations for the selections beyond level four. The 8.0+ prorated levels for levels eight and nine on the A.R.I. are confusing. The numerical designations of the Durrell test selections are not commensurate with the readability levels computed for the selections but the oral and silent selections are commensurate. On the Durrell test, there are no consistent distances between the paragraphs in either subtest when the difference between successive readability levels of the selections are compared.

It may be wise for persons using the Spache scales or the A.R.I. to employ the same readability scales used in the construction of the tests to prepare or select materials thought to be commensurate with the readability levels of the selections on the tests. It is evident that the analysis of the selections utilizing another readability formula resulted in gross descrepancies. However, that suggestion is not meant to imply that students who score at a particular level on either test will be able to read in materials of the same level when the same readability scale is employed. That analysis was beyond the scope of this study. The Durrell test selections will not lend themselves to the procedure recommended above as a result of the different procedures employed to administer the test.

Correlation

High significant correlations were obtained between almost all of the

subtest scores considered for first, second, fourth and sixth grade students. Undoubtedly the rank orders of the students' scores on the two measures obtained for each of the three tests were very similar. All three tests were designed to be administered and scored differently. The readability levels of the selections among the three tests were not commensurate. Undoubtedly, based upon individual standing or position in a group, the students who participated in this study maintained their positions on the six measures computed. Based upon the high correlations obtained, it may be concluded that the individual students' reading abilities remained at the same levels but as a result of using the authors' directions for scoring the students' performance, the examiners view of the proficiency exhibited altered whatever was true about the students' reading ability.

Comparisons

The students' group or means on the subtests were dissimilar enough to warrant the conclusion that the tests, subtests, or various selections from the tests are not interchangeable. Overall results were interpreted to mean that the Durrell test scores were closer to the second, fourth, and sixth grade students' grade placement. Results from the Spache scales were somewhat higher than the students' grade placement. Neither finding is relevant or meaningful as a result of the various connotations of "grade level" and the variability that is known to exist between test scores and test readability levels, and the readability levels of instructional materials found in classrooms and clinics. The important finding related to the grade equivalent mean scores of th- four groups of students was that both the Spache scales mean scores of the students reflected a two year difference between second and fourth, and fourth and sixth grade scores. The Spache scale scores also showed almost a one year difference between the first and second grade students' scores. The Durrell test mean scores did not result in a two year difference between grades, but the differ-

ences were consistent. Apparently both tests may be useful for distinguishing between students' achievement levels and the growth that results from instruction. However, the sensitivity of the tests to growth following instruction was beyond the scope of this study.

The A.R.I. results were not as accurate as those obtained from the other two tests when subjected to only an arithmetical analysis of the distance between the means for the four groups.

Further analysis of the scores revealed that the standard errors of the means (95% confidence level) of the Durrell test scores ranged between .24 and .56 or a possible chance of score fluctuation of 13.4 months; the Spache scale scores ranged between .32 and .72 or a possible chance of score fluctuation of 17.28 months; and the A.R.I. scores ranged between .32 and .96 or a possible chance of score fluctuation of 23.04 months. The readers may draw their own conclusions, or reduce the number of months of chance fluctuations by one half, if they desire to reduce the chance of obtaining an accurate score to fifty percent rather than ninety-five percent. The standard errors of the mean and standard deviations for the Spache scales are similar to those reported in Spache's manual. Normative data for those same comparisons were not available from the other test designers.

Additional Analysis

Authorities other than Spache supported the idea that silent reading and oral reading usually involve different processes. Many also contended that silent reading scores or levels should be higher than those of oral reading, especially for students beginning approximately at third grade level. The data in this report may not be used to support those claims. The scores on neither the Durrell test nor the Spache scales resulted in interpretably higher scores on silent reading when compared to oral reading scores. The theoretical construct upon which those claims are made are sound and supported

by this investigator. However, the students who participated in this study appeared to read silently in the same manner they read orally. Three reasons are hypothesized for that: (1) the students were not taught flexible reading habits, therefore could not apply them in silent reading; (2) testing situations result in mild anxiety which may cause students to revert to close, careful reading that will insure success; and/or (3) students are not given the necessary purposes that provide data that may be used to choose a particular flexible reading rate.

Appendix A contains all the data collected, organized by grade level to allow closer inspection and interpretation by the reader.

BIBLIOGRAPHY

Anderson, W.W. (b) "Focus on Measurement and Evaluation, Commercial Informal Reading Inventories: A Comparative Review," Reading World, 17, (December, 1977), 99-104.

Anderson, W.W. (a) "Focus on Measurement and Evaluation, Informal Reading Inventories: Commercial or Conventional?" Reading World, 17, (October, 1977), 64-70.

†
Durrell, D.D. Durrell Analysis of Reading Difficulty. New York: Harcourt, Brace, and Javanovich, 1955.

Durrell, D.D. Improvement of Basic Reading Abilities. Yonkers-on-Hudson, New York: World Book Company, 1940; pp. 296-315.

Farr, R. Reading: What Can Be Measured? Newark, Delaware: International Reading Association, 1969. pp. 97-109 (a); pp. 85-91 (b).

Goodman, K.S. (Editor) Miscue Analysis: Application to Reading Instruction. Urbana, Illinois: National Council of Teachers of English and ERIC Clearinghouse on Reading and Communications Skills, 1973.

Spache, G.D. "A Comparison of Certain Oral Reading Tests," Journal of Educational Research, 43, (1950), 441-452.

Spache, G.D. Diagnosing and Correcting Reading Disabilities. Boston: Allyn and Bacon, Incorporated, 1976 (c); pp. 154-191.

Spache, G.D. Diagnostic Reading Scales. Monterey, California: California Test Bureau, 1963.

Spache, G.D. Investigating the Issues of Reading Disabilities. Boston: Allyn-Bacon, Incorporated, 1976; pp. 299-319 (a); 152-156 (b).

Woods, M.L. and Moe, A.J. Analytical Reading Inventory. Columbus, Ohio: Charles E. Merrill Publishing Company, 1977.

Bormuth, John R. The cloze readability procedure. In John R. Bormuth (Ed.), Readability in 1968, a research bulletin prepared by a committee of the National Conference on Research in English. Champaign: National Council of Teachers of English, 1968, Pp. 40-47.(a)

Bormuth, John R.(Ed.) Readability in 1968, a research bulletin prepared by a committee of the National Conference on Research in English. Champaign: National Council of Teachers of English, 1968.(b)

Table 1

Readability Levels of the A.R.I., Durrell Test, and Spache Scales As Determined by Using Several Scales

	ARI-FORM A			DURRELL				SPACHE									
				Oral		Silent		A Selections			B Selections			C Selections		D Selections	
	Spache	Harris Jacobson	Fry	Spache	Fry	Spache	Fry	Fry	Spache	Dale-Chall	Fry	Spache	Dale-Chall	Fry	Spache	Fry	Spache
P	1.5	-	.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	1.7	-	2.3	1.5	1.3	1.5	1.3	.8	1.6	-	1.3	1.6	-	.8	1.8	1.8	1.8
2	2.5	-	.8	1.7	1.5	1.8	1.5	1.5	2.3	-	1.8	2.3	-	2.8	2.8	3.3	2.8
3	3.2	-	4.8	2.1	1.8	2.2	3.3	2.5	3.3	-	4.3	3.3	-	3.5	3.8	5.5	3.8
4	-	4.0	4.5	2.8	1.5	2.7	2.3	6.8	-	4.5	6.3	-	4.5	-	-	-	-
5	-	6.0	5.5	3.1	3.8	3.3	3.8	4.3	-	5.5	4.3	-	5.5	-	-	-	-
6	-	7.0	8.5	-	4.8	-	6.8	3.3	-	6.5	6.3	-	6.5	-	-	-	-
7	-	7.0	8.3	-	8.8	-	10.3	7.8	-	7.5	5.8	-	7.5	-	-	-	-
8	-	8.0+	7.5	-	12.+	-	12.+	7.3	-	8.5	8.5	-	8.5	-	-	-	-
9	-	8.0+	7.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 2

Means, Standard Deviations, and Standard Errors of the Mean for Students in Grades One, Two, Four, and Six on the A.R.I., Durrell Test, and Spache Scales

		A.R.I.						DURRELL						SPACHE					
		Oral-Ins			Oral-Ind			Oral			Silent			Oral-Ins			Oral-Ind		
Gr-Mo	N	$\bar{X}$	SD	SE_m	$\bar{X}$	SD	SE_m	$\bar{X}$	SD	SE_m	$\bar{X}$	SD	SEm	$\bar{X}$	SD	SE_m	$\bar{X}$	SD	SE_m
1-9	20	.5	.72	.16	-	-	-	2.56	.61	.13	-	-	-	2.79	.72	.16	-	-	-
2-9	21	2.04	1.65	.36	1.44	1.26	.27	3.11	1.18	.25	3.13	1.31	.28	3.55	1.65	.36	3.47	1.66	.36
4-9	44	5.00	2.20	.33	3.91	2.13	.32	4.89	1.45	.21	4.76	1.44	.21	6.25	1.84	.27	6.22	2.14	.32
6-9	52	8.31	1.44	.36	6.81	1.93	.48	6.47	.49	.12	6.15	.70	.17	8.0	.89	.22	8.06	1.09	.27

Table 3

Differences in Mean Scores For

First, Second, Fourth

And Sixth Grade Students On The

A.R.I., Durrell Test, And Spache Scales

	A.R.I.		DURRELL		SPACHE	
	Oral Ins	Oral Ind	Oral -	Silent -	Oral Inst	Silent Ind
1-0	.50	-	2.56	-	2.79	-
2-1	1.54	1.44	.55	-	.76	-
4-2	2.96	2.47	1.78	1.63	2.10	2.15
6-4	3.31	2.90	1.58	1.39	1.75	1.84

Table 4

Pearson Product-Moment Correlations

Between Student Subtest Scores, Grades

One, Two, Four, and Six on the A.R.I.

Durrell Test, And Spache Scales

		GRADE ONE		GRADE TWO				GRADE FOUR				GRADE SIX			
		DURRELL	SPACHE INST	DURRELL		SPACHE INST	SPACHE IND	DURRELL		SPACHE INST	SPACHE IND	DURRELL		SPACHE INST	SPACHE IND
		Oral	Oral	Oral	Silent	Oral	Silent	Oral	Silent	Oral	Silent	Oral	Silent	Oral	Silent
A. R. I.	Oral Ind			.938	.620	.918	.726	.992	.864	.977	.861	.939	.783	.902	.587
	Oral Ins	.932	,917	.964	.936	.950	.980	.995	.995	.991	.992	.896	.868	.958	.982
DURRELL	Oral		.990			.986				.990				.925	
	Silent						.989				.989				.897
		N=20		N=21				N=44				N=32			
		P.01=.561		P.01=.548				P.01=.376				P.01=.448			

Table 5

Correlated t Test Results

In Which No Significant Differences

Were Shown Between Subtest Scores

On The A.R.I., Durrell Test, and Spache Scales

Test			A.R.I.					DURRELL		
LEVELS			INSTRUCTIONAL			INDEPENDENT				
PERFORMANCES			ORAL			ORAL		ORAL		SILENT
		GR.	2	4	6	4	6	1	2	2
DURRELL	ORAL	2	2.39	-	-	-	-	-	-	-
		4	-	.284	-	2.51	-	-	-	-
		6	-	-	-	-	.674	-	-	-
	SILENT	2	2.36	-	-	-	-	-	-	-
		4	-	-	-	2.17	-	-	-	-
		6	-	-	-	-	1.27	-	-	-
SPACHE	ORAL	1	-	-	-	-	-	1.04	-	-
		2	-	-	-	-	-	-	.996	-
		6	-	-	.734	-	2.22	-	-	-
	SILENT	2	-	-	-	-	-	-	-	.730
		4	-	2.50	-	-	-	-	-	-
		6	-	-	.551	-	2.24	-	-	-
P.01			2.70	2.62	2.64	2.62	2.64	2.71	2.70	2.70
N			21	44	32	44	32	20	21	21

APPENDIX A

GRADE ONE N = 20

	DURRELL ARI ORAL (inst)	SPACHE ARI ORAL (inst)	DURRELL SPACHE ORAL	DF	P.01
r	.932	.917	.99	18	.561
t	9.603	9.901	1.04	38	2.713

	DURRELL (0)	SPACHE (0)	ARI (0)
$\bar{x}$	2.56	2.79	.53
S.D.	.61	.72	.72
$SE_{\bar{x}}$	.13	.16	.16

GRADE TWO N = 21

	DURRELL ARI/IND ORAL	SPACHE ARI/IND ORAL	SPACHE DURRELL ORAL	DF	P.01
r	.938	.918	.986	19	.584
t	4.411	4.644	.996	40	2.704

	DURRELL (0)	SPACHE (0)	ARI (0-ind)
$\bar{x}$	3.11	1.55	1.44
S.D.	1.18	1.65	1.26
$SE_{\bar{x}}$	.25	.36	.27

GRADE TWO N = 21

	DURRELL ARI/INST ORAL	SPACHE ARI/INST ORAL	SPACHE DURRELL ORAL	DF	P.01
r	.964	.950	.986	19	.548
t	2.39	2.952	.996	40	2.704

	DURRELL (0)	SPACHE (0)	ARI (0-inst)
$\bar{x}$	3.11	3.55	2.04
S.D.	1.18	1.65	1.65
$SE_{\bar{x}}$	.25	.36	.36

GRADE TWO N=21

	DURRELL (SILENT) ARI/INST. (ORAL)	SPACHE (SILENT) ARI/INST. (ORAL)	SPACHE (SILENT) DURRELL (SILENT)	DF	P.01
r	.963	.980	.989	19	.548
t	2.362	2.787	.730	40	2.704

	DURRELL (S)	SPACHE (S)	ARI (0-inst)
$\bar{x}$	3.13	3.47	2.04
S.D.	1.31	1.66	1.65
$SE_{\bar{x}}$	.28	.36	.36

GRADE TWO N=21

	DURRELL (SILENT) ARI/IND (ORAL)	SPACHE (SILENT) ARI/IND (ORAL)	SPACHE (SILENT) DURRELL (SILENT)	DF	P.01
r	.620	.762	.989	19	.548
t	4.222	4.451	.730	40	2.704

	DURRELL (S)	SPACHE (S)	ARI (0-inst)
$\bar{x}$	3.13	3.47	1.44
S.D.	1.31	1.66	1.26
$SE_{\bar{x}}$	.28	.36	.27

GRADE FOUR N=44

	DURRELL ARI/IND ORAL	SPACHE ARI/IND ORAL	SPACHE DURRELL ORAL	DF	P.01
r	.992	.977	.990	42	.376
t	2.515	5.489	3.827	86	2.625

	DURRELL (O)	SPACHE(O)	ARI (O-ind)
$\bar{x}$	4.89	6.25	3.91
S.D.	1.45	1.84	2.13
$SE_{\bar{x}}$	.21	.27	.32

GRADE FOUR N=44

	DURRELL ARI/INST ORAL	SPACHE ARI/INST ORAL	SPACHE DURRELL ORAL	DF	P.01
r	.995	.991	.990	42	.376
t	.284	2.865	3.827	86	2.625

	DURRELL (0)	SPACHE (0)	ARI (0-inst)
$\bar{x}$	4.89	6.25	5.00
S.D.	1.45	1.84	2.20
$SE_{\bar{x}}$	.21	.27	.33

GRADE FOUR N=44

	DURRELL (S) ARI/INST (0)	SPACHE (S) ARI/INST (0)	SPACHE (S) DURRELL (S)	DF	P.01
r	.995	.992	.989	42	.376
t	.616	2.509	3.742	86	2.625

	DURRELL (S)	SPACHE (S)	ARI (0-inst)
$\bar{x}$	4.76	6.22	5.00
S.D.	1.44	2.14	2.20
$SE_{\bar{x}}$	.21	.32	.33

GRADE FOUR N=44

	DURRELL (S) ARI/IND (0)	SPACHE (S) ARI/IND (0)	SPACHE (S) DURRELL (S)	DF	P.01
r	.864	.861	.989	42	.376
t	2.176	5.045	3.742	86	2.625

	DURRELL (S)	SPACHE (S)	ARI (0-ind)
$\bar{x}$	4.76	6.22	3.91
S.D.	1.44	2.14	2.13
$SE_{\bar{x}}$	.21	.32	.32

GRADE SIX N=32

	DURRELL ARI (IND) ORAL	SPACHE ARI/IND ORAL	SPACHE DURRELL ORAL	DF	P.01
r	.939	.902	.925	30	.448
t	.674	2.223	5.968	62	2.642

	DURRELL(0)	SPACHE (0)	ARI (0-ind)
$\bar{x}$	6.47	8.00	6.81
S.D.	.49	.89	1.93
$SE_{\bar{x}}$	.12	.22	.48

GRADE SIX N=32

	DURRELL ARI/INS ORAL	SPACHE ARI/INST ORAL	SPACHE DURRELL ORAL	DF	P.01
r	.896	.958	.925	30	.448
t	4.804	.734	5.96	62	2.642

	DURRELL (0)	SPACHE (0)	ARI (0-inst)
$\bar{x}$	6.47	8.00	8.31
S.D.	.49	.89	1.44
$SE_{\bar{x}}$	.12	.22	.36

GRADE SIX N=32

	DURRELL (S) ARI/INST (0)	SPACHE (S) ARI/INST (O)	SPACHE (S) DURRELL (S)	DF	P.01
r	.868	.982	.897	30	.448
t	5.363	.551	5.871	62	2.642

	DURRELL (S)	SPACHE(S)	ARI (O-ins)
$\bar{x}$	6.15	8.06	8.31
S.D.	.70	1.09	1.44
$SE_{\bar{x}}$	.17	.27	.36

GRADE Six N=32

	DURRELL (S) ARI/IND (O)	SPACHE (S) ARI/IND (O)	SPACHE (S) DURRELL (S)	DF	P.01
r	.783	.587	.897	30	.448
t	1.272	2.245	5.871	62	2.642

	DURRELL (S)	SPACHE (S)	ARI (O-ind)
$\bar{x}$	6.15	8.06	6.81
S.D.	.70	1.09	1.93
$SE_{\bar{x}}$	.17	.27	.48